

## **VERIFICATION**

I, Mamoru FUKUSHIMA , a national of Japan, NITTO IPO Ltd.,  
8TH Floor, NO. 17 ARAI BLDG., 1-3-3, Shinkawa, Chuo-ku, Tokyo  
104-0033, Japan, verify that to the best of my knowledge and belief the  
following is a true translation made by me of the annexed document  
which is Japanese Application, No. Hei 10-328940 filed on November 19,  
1998.

Dated this 28th day of January, 2004

  
Mamoru Fukushima  
Mamoru FUKUSHIMA, Translator

[Document Name] Application for Patent

[Reference No.] PNT980870

[Application Date] November 19, 1998

[Destination] Commissioner, Patent Office

[International Patent Classification] G06F 17/00

[Title of the Invention] A method of searching documents  
and a service for searching documents

[No. of Claims] 7

[Inventor]

[Address] 2520, Akanuma, Hatoyama-machi,  
Hiki-gun, Saitama-ken  
Hitachi, Ltd.,  
Advanced Research Laboratory

[Name] Makoto IWAYAMA

[Inventor]

[Address] 2520, Akanuma, Hatoyama-machi,  
Hiki-gun, Saitama-ken  
Hitachi, Ltd.,  
Advanced Research Laboratory

[Name] Shingo NISHIOKA

[Inventor]

[Address] 2520, Akanuma, Hatoyama-machi,  
Hiki-gun, Saitama-ken  
Hitachi, Ltd.,  
Advanced Research Laboratory

[Name] Yoshiki NIWA

[Inventor]

[Address] 2520, Akanuma, Hatoyama-machi,  
Hiki-gun, Saitama-ken  
Hitachi, Ltd.,  
Advanced Research Laboratory

[Name] Akihiko TAKANO

[Applicant for Patent]

[Identification No.] 000005108  
[Name] Hitachi, Ltd.

[Agent]

[Identification No.] 100061893  
[Patent Attorney] Akio TAKAHASHI  
[Telephone No.] 03-3661-00071

[Agent]

[Identification No.] 100086656  
[Patent Attorney] Kyosuke TANAKA  
[Telephone No.] 03-3661-00071

[Designation of Charge]

[Ledger No. for Prepayment] 011626  
[Amount of Payment] 21000

[List of Document Submitted]

[Object Name] Specification 01  
[Object Name] Drawings 01  
[Object Name] Abstract 01

[Proof] Required

**【NAME OF DOCUMENT】**      SPECIFICATION

**【TITLE OF THE INVENTION】**

A method of searching documents and a service for  
searching documents

5      **【WHAT IS CLAIMED IS】**

**【Claim 1】**

A document search method having a function to change  
over between plural document databases, and a function  
to search a set of documents having a high

10     relevance to a search input from a selected document  
database in the order of higher relevance, this input  
being a set of keywords, fragments of a document or any  
desired set of documents, wherein the search results  
from said document database can be used as input for  
15     searching another database.

**【Claim 2】**

A document search method as defined in Claim 1,  
wherein an interface is provided in which a set of  
documents from the search result of one document  
20     database can be selected or deselected, and a set of  
documents selected from the search result can be used  
as input to perform a search on another database.

**【Claim 3】**

A document search method as defined in Claim 1,  
25     wherein a summary containing only topic words in the  
search input is used to perform a search.

**【Claim 4】**

A document search method as defined in Claim 1,

wherein servers comprising document databases and programs to manipulate said databases are dispersed over a network, a client transmits a set of documents in a search input to a server where a selected document database is stored, receives a summary comprising only topic words related to the set of documents which is sent, sends a search input corresponding to said summary reflecting a user's evaluation of the summary to a server where another document database is stored, and receives a search result.

**[Claim 5]**

A document search method as defined in Claim 4, wherein said server produces a summary from topic words relevant to a set of documents sent by the client and transmits it to the client, and searches and transmits a set of documents having a high relevance to any summary sent by the client, to the client.

**[Claim 6]**

A document search method as defined in Claim 4, wherein said client has an interface for specifying a set of documents for search input and document databases to be searched, the set of documents in the search input is sent to a server specified by the user, a summary of the set of documents is received from this server, the summary received is sent to a server comprising another document database, and search results are received from the latter server and displayed.

**【Claim 7】**

A service for searching documents wherein servers comprising document databases and programs to manipulate said databases are dispersed over a network.

5 and a client connected to said servers performs a document search, said service providing a function for the client to transmit a set of documents in a search input to one of said servers where a selected document database is stored, receive a summary comprising only

10 topic words related to the set of documents which is sent, send a search input corresponding to said summary reflecting a user's evaluation of the summary to a server where another document database is stored, and receive a search result, wherein said server produces a

15 summary of topic words relevant to the set of documents sent by the client and transmits it to the client, and searches and transmits a set of documents having a high relevance to any desired summary sent by the client, to the client.

20

**【DETAILED DESCRIPTION OF THE INVENTION】**

**【0001】**

**【Technical Field of the Invention】**

This invention relates to a document searching method for changing over between plural document databases, and constructing relationships between plural document databases.

25

**【0002】**

**【Prior Art】**

As more and more document information is converted to electronic format, a greater need is emerging to search different types of document database simultaneously. For instance, users often wish to look up dictionaries relating to newspaper articles which they may find of interest.

5           【0003】

In the past, it was possible to perform a search independently by changing over between plural document databases, but there was no way of examining the relevancy of sets of documents in other databases to a set of documents in one particular database.

10           【0004】

If however the search is limited to the same document database, it is possible to search other document sets within that database. In this case, sufficient search speed is often obtained by calculating the relevance between documents before searching. Even with different databases, it is possible to search plural document databases at the same time if such a calculation is performed beforehand, but since the need for calculation increases as the number of databases increases due to increasing numbers of combinations, this method is not realistic.

20           【0005】

It is also possible to first analyze the set of key documents on the user side to compose a search input, and then search in other document databases by

using the input, but in this case, the user side has to receive all the information about the set of key documents, and if the document databases are on a network, the amount of traffic would be huge.

5           **【0006】**

**【Problems to be Solved by the Invention】**

It is therefore an object of this invention to resolve the problems inherent in existing technology by allowing a user to specify an arbitrary set of 10 documents in an arbitrary document database, and to efficiently search sets of documents relating to this set of documents from within any particular database.

**【0007】**

**【Means to Solve the Problems】**

15           When there is a large search input as in the case of a set of key documents, instead of using all the information in the search input, it is faster to perform a search using only topic words of the search input as a summary, and this also reduces the load on 20 the network. In the context of this specification, "summary" means a "set of topic words for a set of documents".

**【0008】**

The document databases are located on servers 25 on a network comprising a module for building a summary by selecting topic words for a set of documents within the document database, and a module for performing a search on any arbitrary summary.

**【0009】**

- A user who performs a search specifies a set of documents via a client to a server in which an source document database is stored, and receives a summary.
- 5 Next, the summary is sent to a server where a target document database to be searched is located, and a search result is received.

**【0010】**

- As the search interface of the client, a display area for a set of documents is first provided wherein the required set of key documents can be specified, and the database to be searched can also be selected. In the client, the user then selects an interesting set of documents from among a set of documents displayed in this display area, and if necessary, changes over the document database which is to be searched.
- 10  
15

**【0011】**

**【Preferred Embodiments of the Invention】**

20 Description of the Preferred Embodiments

Fig. 1 shows a typical general arrangement wherein a client 11 specifies an arbitrary set of key documents in a document database 131 of a server 13, and obtains a set of documents having a high relevance (similarity) to the specified set of key documents from a document database 141 of another server 14. Herein, the source and target document databases 131, 141 are located on servers in different places which can be

25

respectively accessed via a network 12.

**【0012】**

First, the client 11 specifies a set of key documents in the source document database 131 according 5 to user's specification, and sends this information to the server 13 as a set comprising a document identifier, for example an ID attached to each document which the server 13 can understand, via the network 12. The set of documents is specified in a window for displaying 10 search results P1 described later.

**【0013】**

The server 13 identifies a set of documents which were sent from the client. A summary of the set of documents is then made for the searched set of 15 documents by a summary making module 132, and this is sent back to the client 11 via the network 12. Herein, the term "summary" means a set of topic words relating to a set of documents. The summary making module can be constructed by any of the known methods such as that 20 disclosed in Japanese Patent Laid-Open No. Hei 9-62693, "Method of Document Classification by Probability Model".

**【0014】**

As an example, word frequencies are first 25 totaled by splitting up all the documents in the set of documents for which it is desired to make a summary, into words. In general, as the degree to which a given set of documents is represented by particular words is

higher for words which appear more frequently within it, words will tend to be included in the summary more frequently the higher their occurrence frequency is in the set of documents. However, general words which often appear in all documents such as "do", etc., are not suitable for the summary. Therefore, words are usually selected for inclusion in the summary by considering also their appearance frequency in the document database to which the set of documents belongs.

5 Specifically, desirable words are topic words having a high occurrence frequency in a specified set of documents but a low overall frequency in the document database overall, i.e., they are suitable for a summary characterizing the set of documents. Hence, words are

10 selected for the summary by calculating their weighting from suitable parameters using the occurrence frequency in the set of documents and the occurrence frequency in the document database as input, and adopting words having a weighting equal to or greater than a certain

15 threshold. The higher the weighting, the higher the relevance of the word to a given document, and the lower the weighting, the lower the relevance of the word to the document. The server 13 returns a set of words having a weighting calculated by the above-

20 mentioned method to the client via the network 12.

25 These words are displayed as "topic words" in Fig. 2.

【0015】

Next, at the client 11, users evaluate or

process the summary (summary of the set of key documents) from the server 13, and the client 11 transmits it to the target server 14 via the network 12.

**【0016】**

5        In the evaluation or the processing performed by users at the client, users for example remove words which are not deemed to be relevant from the summary, or users for example replace words in the summary.  
Using the search module 143, the server 14 calculates  
10      the relevance of the summary of the set of key documents sent from the client to the target document database 141, and returns document identifiers of high relevance to the client 11 with a relevance weighting.  
The search module here can be implemented by a keyword  
15      search known in the art. Specifically, as the summary of the set of documents which is input is a set of words with weightings, these words may be considered as weighted input keywords and an OR keyword search performed. In this case, the weighting (relevance) of  
20      the document which is the search result can be calculated. This is done by taking words which appear both in the summary and the document to be searched, calculating an overall weighting from their weighting in the summary and their weighting in the document to be searched (e.g., product of the two weightings), and then adding up the weightings of all the words (e.g., calculating a sum total) to obtain the relevance.  
25

**【0017】**

Using the above method, the client 11 can obtain a set of documents in the document database 141 which relates to an arbitrary set of key documents in the document database 131. The characteristic feature 5 of this method is that network traffic is reduced to a small amount by leaving the processing (summary making) of the original set of documents searched to the server side. It will be appreciated that the amount of traffic is much less than in the case where the client 10 has to receive and process all of the document information to be searched. The search assistant module 112 of the client then basically has only to send the summary of the set of documents from the source server to the target server, and almost all of 15 the processing involved in the search can be left to both servers. Moreover, the server side merely has to have a summary making module and a search module for the document database in question, and it is therefore completely unnecessary to consider information in other 20 document databases.

**【0018】**

In the aforesaid description, a method was described wherein the document database 131 was the source database and the document database 141 was the 25 target database, but the same method can be adopted wherein the document database 141 is the source database and the document database 131 is the target database. In this case, the client obtains a summary

of the set of key documents from a summary making module 142 of the server 14, transmits it to the server 13 which is to be searched, and obtains relevant documents in the document database 131 from the search 5 module 133 of the server 13. If the above is generalized, and a server with a summary making module and a search module is provided for a new document database, this document database can be made to function as the source database or target database for 10 all document databases connected to the network simply by connecting the server to the network.

【0019】

Finally, Fig. 2 shows an embodiment concerning the client. 111 is an example of a search assistant interface installed in the client. This is basically the same as the interface proposed by the inventor of the present application in Japanese Patent Laid-Open No. Hei 11-85786, "Document search support method and 15 document search support service" (corresponding to U.S. Patent Application S.N.09/145,155, filed 09/01/98 by Nishioka et al), or Japanese Patent Laid-Open Hei 10-74210, "Document search support method and device, and document search service using same", (corresponding to 20 U.S. Patent Application S.N.08-888,017, filed 07/03/97 by Niwa et al). E1 is a window for inputting a search query, wherein the user can input a search query by a string of keywords or in the form of a sentence. M1 is 25 a window for selecting a document database wherein the

user can pull down a specific part on the right edge  
with a mouse to show a list of document databases, and  
select a desired document database. B1 is a search  
button which initiates a search. Therefore, the user  
5 inputs an arbitrary search query in the window E1,  
selects a document database to be searched in the  
window M1, and performs an ordinary search by keywords  
input to the window E1 concerning the document database  
selected in the window M1, by pressing the button B1.  
10 This search is performed with the support of the search  
assistant module 112 shown in Fig. 1, but as the  
details of the search method were given in the previous  
application, they will not be repeated here.

【0020】

15 P1 is a window for displaying a search result.  
In the upper part, a panel showing the total number of  
documents retrieved as a result of the search process  
and a number of documents selected by the user as  
described hereafter, is displayed. Underneath this, a  
20 panel is provided for the user to specify selected/not  
selected(P13), and a document title part showing the  
relevance(P12) to the search query and titles(P11) of  
documents displayed in the form of a list. This  
display window has a scroll function so that, by  
25 scrolling, the user can see a part which cannot be  
displayed in the display at one time. In the  
selected/not selected panel, documents are either  
selected or deselected each time there is a mouse click.

When documents are selected by clicking, a summary of the corresponding documents is displayed as a graphical representation of a set of words with weightings in a summary display window P2. The summary display window 5 P2 also has a panel in its upper part where the total number of topic words and the number of topic words selected by the user are displayed. Document titles are usually sorted in order of relevance.

【0021】

10 The window P1 for displaying the search result in the diagram shows that a total of 22 documents were retrieved as a result of the search, and that three documents were selected by the user as interesting documents judging from their titles. The selected 15 documents are given a check mark by clicking. In the summary display window P2, five topic words are accordingly displayed corresponding to the selected documents.

【0022】

20 Although omitted from this embodiment, conversely, documents for which the topic words selected in the summary display window P2 are representative, can be displayed in the window P1. Therefore, the user can perform a more advanced search 25 by making a summary customized according to his preference. This is explained in detail in the aforesaid reference Japanese patent application Hei 9-240963.

**【0022】**

Hence, the user can select/deselect documents while referring to the titles and the topic words of selected documents, and can select plural documents in  
5 which he is interested.

**【0024】**

Subsequently, if the user is interested in handling other document data for the set of documents corresponding to this search result, he may change the  
10 document database in the window M1, and press the button B1 to begin a new search.

**【0025】**

Hence, the client sends an identifier of the plural documents selected to the server where the  
15 source document database is stored (for example, the server 13), obtains a summary of these plural documents, sends this summary to the server where the target document database is stored (for example, the server 14), and obtains a search result from the target server  
20 (for example, the server 14). The new search result is displayed in the window P1. In other words, in this case, P1 is updated by the set of documents which was newly searched.

**【0026】**

25 To compare a new search result with a previous search result, the user may press a back button B2 to re-display the previous search result in the window P1, and return the display of the window P1 to its state

before search was performed. Likewise, the window P1 can be advanced to the new search result by pushing a forward button B3.

**【0027】**

5 As the user can search other document databases corresponding to such a search result at any stage of the search, the user can freely proceed from one database to another database by repeating the search cycle. Naturally, it is also possible to repeat this  
10 cycle within the same document database, i.e., without changing the document database.

**【0028】**

**【Effect of the Invention】**

According to this invention, the user can  
15 freely specify a document database to be searched and freely enhance the search without concern for the location or composition of each document database. Further, as a server in which a document database is located can be modularized, the server can be made to  
20 function as an source database or a target database with respect to all other databases connected to a network simply by connecting a server comprising a summary making module and a search module to the network when it is desired to search new document  
25 databases.

**【BRIEF DESCRIPTION OF THE DRAWINGS】**

**【Fig. 1】**

Fig. 1 is a diagram showing an example of the

overall construction of a system implementing the plural document database search method.

**【Fig. 2】**

5 Fig. 2 is a diagram showing an example of the construction of a search assistant interface in a client.

**【Description of Reference Numbers】**

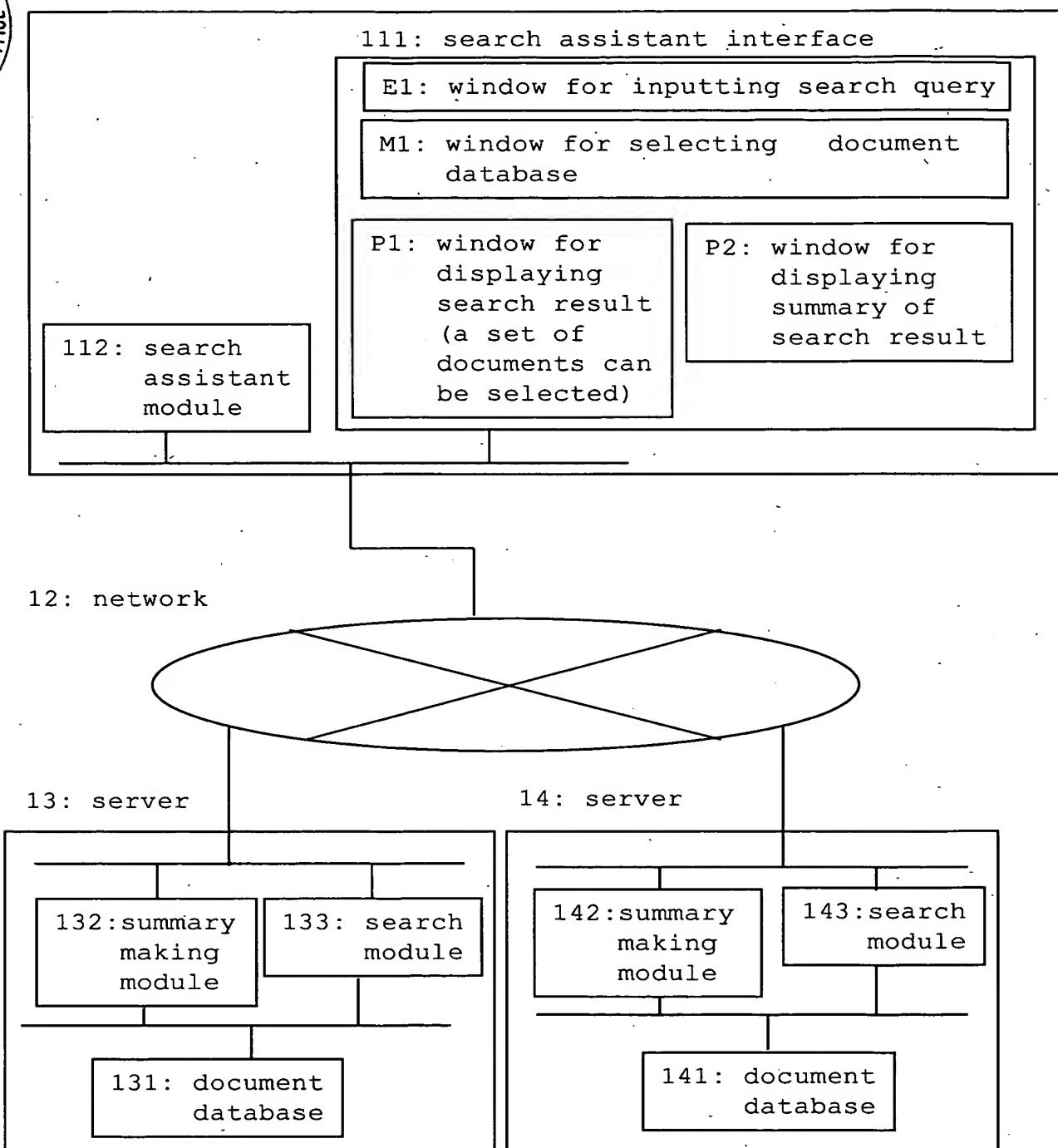
11:client, 111:search assistant interface, 112:search assistant module, 12:network, 13:server, 131:document  
10 database, 132:summary making module, 133: search module,  
14:server, 141:document database, 142:summary making module, 143:search module, B1:search button, B2:back button. B3:forward button, E1>window for inputting search query, M1>window for selecting document database,  
15 P1>window for displaying search result, P2>window for displaying summary of search result.

**【NAME OF DOCUMENT】      DRAWINGS**

【Fig. 1】

11: Client

Figure 1



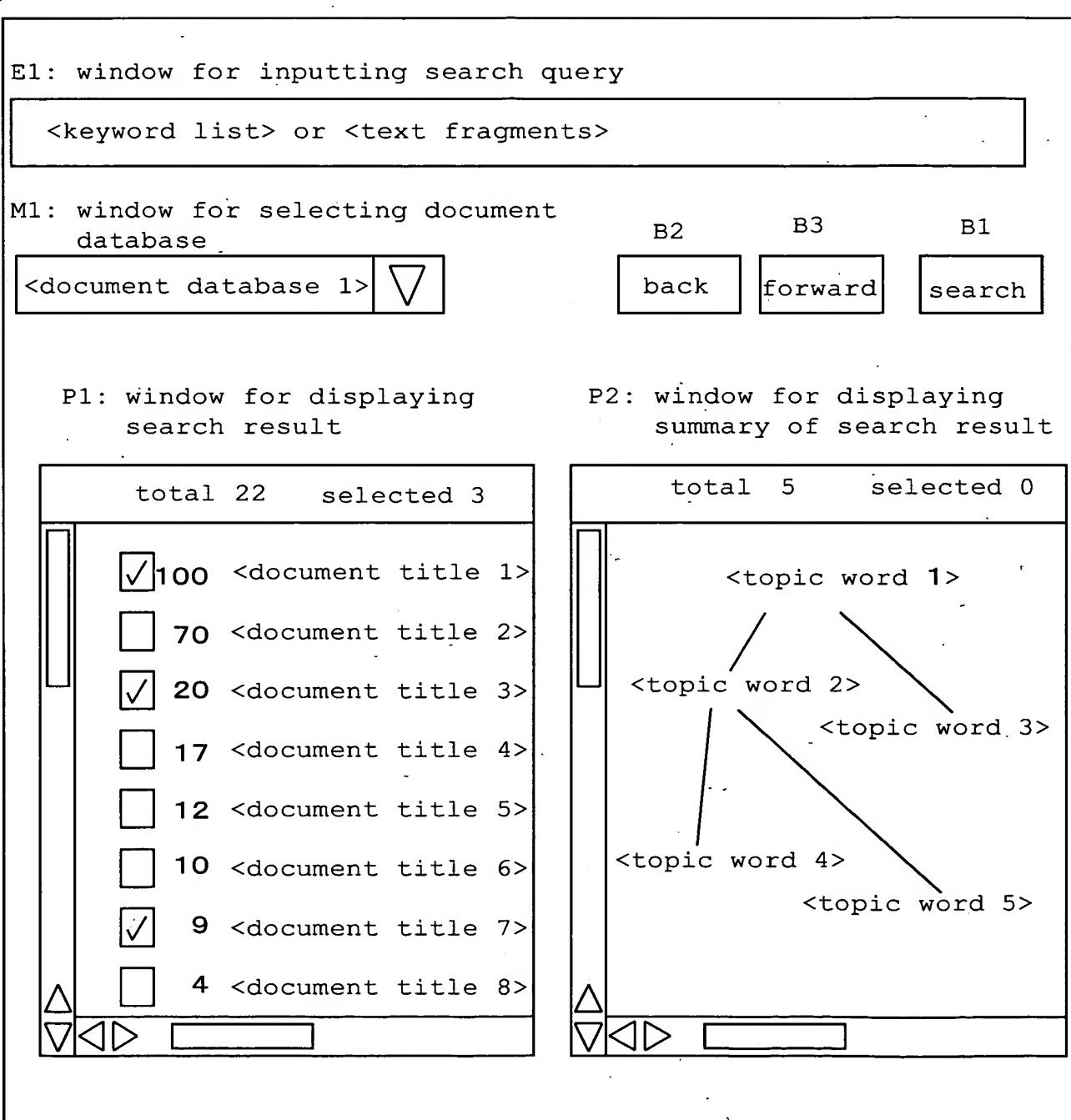
SIPE JCITY  
MAR 23 2004  
SEARCH OFFICE

COPIE JC177  
MAR 23 2004  
SEARCH OFFICE

【Fig. 2】

Figure 2

111: search assistant interface



**【NAME OF DOCUMENT】** ABSTRACT

**【ABSTRACT】**

**【OBJECT】** To provide an efficient means of performing a document search wherein relevance between plural document databases is examined.

5           **【ACHIEVEMENT MEANS】** A summary making module and a search module are provided to document databases, and these are connected to a network as a server. A client obtains a relevant set of documents in a specified 10 document database via this summary from a set of documents in the specified document database. The summary obtained is sent to another server, and a search is performed according to the summary in a document database in the server to which the summary is 15 transferred.

**【SELECTED DRAWING】** Fig. 1